心理科学进展 2018, Vol. 26, No. 11, 1935–1951
Advances in Psychological Science DOI: 10.3724/SP.J.1042.2018.01935

·研究前沿(Regular Articles)·

McGurk 效应的影响因素与神经基础*

罗霄骁 [1] 康冠兰 [1] 周晓林 [1,2,3,4]
([1] 北京大学心理与认知科学学院, 北京 100871)
([2] 北京大学机器感知与智能教育部重点实验室, 北京 100871)
([3] 北京大学 IDG 麦戈文脑科学研究所, 北京 100871)
([4] 浙江师范大学心理与脑科学研究院, 金华 321004)

摘 要 McGurk 效应(麦格克效应)是典型的视听整合现象, 该效应受到刺激的物理特征、注意分配、个体视听信息依赖程度、视听整合能力、语言文化差异的影响。引发 McGurk 效应的关键视觉信息主要来自说话者的嘴部区域。产生 McGurk 效应的认知过程包含早期的视听整合(与颞上皮层有关)以及晚期的视听不一致冲突(与额下皮层有关)。未来研究应关注面孔社会信息对 McGurk 效应的影响, McGurk 效应中单通道信息加工与视听整合的关系, 结合计算模型探讨其认知神经机制等。

关键词 McGurk 效应; 视听言语感知; 视听整合; 多感觉整合
分类号 B842

1 前言

多感觉整合(multisensory integration)是将不同感觉通道输入的信息有效合并为统一、连贯、稳定的知觉的过程(Stein & Stanford, 2008; 文小辉, 李国强, 刘强, 2011; 文小辉等, 2009)。视听言语感知(audiovisual speech perception)是一种典型的多感觉整合过程——在与他人面对面交流时, 个体会整合视觉信息和听觉信息进行言语理解, 也即视听整合(audiovisual integration)。其中, “视觉信息”指的是说话人的口唇发音动作、面部肌肉活动及表情等。个体可以利用这些信息形成连续的视知觉, 并与头脑中储存的词语表象相比较和联系, 进而理解说话者表达的内容。该过程也称为“唇读” (lipreading) (Summerfield, 1992; 朴永馨, 2006; 徐诚, 2013)。例如: 听力障碍者主要依赖视觉信息进行言语感知(雷江华, 方俊明, 2005)。“听觉信息”指的是说话人的语音信息。对听力正常者而言, 听觉信息在言语感知中起主导作用, 视觉信息是辅助信息。即使如此, 视觉信息对言语感知的影响仍然存在, 例如同时呈现听觉信息和相应的视觉信息时, 言语感知准确率比单独呈现听觉信息时高(Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007)——这体现了视听整合的益处。

McGurk 效应(McGurk effect / McGurk illusion) (McGurk & MacDonald, 1976)是一种典型的视听整合现象, 指的是当特定发音的视觉刺激与特定发音的听觉刺激同时呈现时, 个体可能产生新感知的现象(例如: 说话者说“ga”的视频和说“ba”的音频同时呈现, 听话者可能会感知到另一个音节“da”), 这反映了视觉信息对听觉感知的影响。一般认为, 发生了 McGurk 效应即发生了视听整合, 所以 McGurk 效应发生率可以作为视听整合强弱的指标(Fernández, Macaluso, & Soto-Faraco, 2017; Marques, Lapenta, Costa, & Boggio, 2016; Tiippana, 2014)。

McGurk 效应一直是视听言语感知研究中的热点问题。自 McGurk 和 MacDonald (1976)发表该效应, 到 2016 年 40 年间, 原文已经被引用近

收稿日期: 2018-03-13
* 国家自然科学基金面上项目(31470976), 科技部 973 项目(2015CB856400), 机器感知与智能教育部重点实验室开放课题基金项目(K-2017-05)。
通信作者: 周晓林, E-mail: xz104@pku.edu.cn

5000 次(Alsius, Paré, & Munhall, 2018; MacDonald, 2018)。即使如此，目前仍然缺乏全面、系统的 McGurk 效应综述。Marques 等人(2016)的综述主要关注 McGurk 效应的研究对理解视听整合过程的启示，尤其是如何用视听整合的理论模型来解释 McGurk 效应，以及 McGurk 效应在特殊人群言语感知研究中的应用。但该综述集中于视听整合问题，对 McGurk 效应本身关注不足。例如：没有关注 McGurk 效应的测量和界定；对 McGurk 效应的影响因素讨论较少(文中只涉及了外界物理刺激的影响)；没有关注 McGurk 效应中可能存在的视听不一致冲突问题等。Alsius 等人(2018)的综述主要关注 McGurk 效应作为视听言语感知过程的研究工具有哪些局限性以及需要注意的问题，尤其是影响 McGurk 效应发生率的因素以及 McGurk 刺激与视听一致刺激的差异。但该综述的主要目的在于反思当前研究使用 McGurk 范式的合理性，没有涉及神经基础问题；且其对 McGurk 效应的影响因素的阐述系统性不足。MacDonald (2018)的综述回顾了 40 年前 McGurk 效应的发现过程以及作者的心路历程，是对历史事件的回顾，没有关注 McGurk 效应的最新研究进展。

本文尝试对 McGurk 效应进行全面、系统的综述。首先探讨 McGurk 效应的测量与界定问题。再从个体内变异和个体间变异的角度出发，阐述影响 McGurk 效应的相关因素。进一步从眼动模式、动态神经加工过程、相关脑区三个方面，阐述 McGurk 效应的认知神经基础。最后提出未来研究展望以及需要注意的问题。

2 McGurk 效应的测量与界定

已有研究一般采用“McGurk 效应发生率”作为评价 McGurk 效应强弱(多少)的指标——使用 McGurk 刺激实施多次测量后，计算其中发生 McGurk 效应的次数比例(在测量的过程中需要加入视听一致刺激或视听不一致但不会诱发 McGurk 效应的刺激作为填充试次)。研究中最常用的 McGurk 刺激是视觉“ga”加听觉“ba”的视听组合，发生 McGurk 效应时可能感知到“da” (Beauchamp, Nath, & Pasalar, 2010; Fernández et al., 2017; Nath & Beauchamp, 2012)。除此之外，视觉“ka”加听觉“pa”可能感知到“ta” (Gurler, Doyle, Walker, Magnotti, & Beauchamp, 2015)。另一方面，也有研究在元音上采用“i”等其他搭配，例如视觉“gi”加听觉“bi”可能感知到“di” (Colin, Radeau, Soquet, Demolin, Colin, & Deltenre, 2002)。也有研究在辅音之前添加元音，例如视觉“aga”加听觉“aba”可能感知到“ada” (Bertelson, Vroomen, & de Gelder, 2003; Buchan & Munhall, 2012)。还有研究会重复两次音节，例如视觉“gaga”加听觉“baba”可能感知到“dada” (Mallick, Magnotti, & Beauchamp, 2015; McGurk & MacDonald, 1976)。虽然 McGurk 刺激有很多种，但是其核心都是特定视觉辅音和听觉辅音的组合，使个体感知到的听觉刺激发生改变。

为什么只有特定的视听信息组合才会产生 McGurk 效应，而其他组合则不会？分层预测编码模型(hierarchical predictive coding model, Olasagasti, Bouton, & Giraud, 2015)给出了解释。该模型考虑到视觉和听觉信息的动态交互过程，建立了视觉信息(唇形，lip aperture)和听觉信息(第二共振峰，second formant)在物理维度上的动态变化二维空间，以探究不同感觉通道对输入的感知信息进行预测和判断的动态变化过程。在典型的 McGurk 效应中，视觉“ga”和听觉“ba”的视听不一致输入与视觉“da”和听觉“da”的视听一致输入在上述二维空间中的坐标非常接近，所以这种情况下的视听不一致并不会造成很强的跨通道冲突，而可能会更接近“da”的表征。但如果反过来，对于视觉“ba”和听觉“ga”的不一致输入，其坐标与其他视听一致音节的坐标都不接近，因此，这种视听不一致信息输入会造成较强跨通道冲突，无法融合。所以，融合的发生可能是因为视听不一致刺激的视听通道表征在二维动态编码空间中非常接近某个视听一致刺激的表征，大脑就更容易预期当前刺激是视听一致的，进而表征出在二维空间中坐标接近的视听一致感知。

关于 McGurk 效应的界定，即“被试的什么反应可以算作发生了 McGurk 效应”，不同的研究之间存在一定差异。部分研究的界定比较严格——只有个体感知到了特定的融合音节(例如在呈现视觉“ga”和听觉“ba”时感知到“da”)，才能算是发生了 McGurk 效应(Colin et al., 2002; Rosenblum, Schmuckler, & Johnson, 1997)。但是这种界定方式忽略了很多其他情况(例如：依据该界定，如果被试报告感知到“tha”“ga”等其他音节，就不能算作发生了 McGurk 效应)。所以，另一部分研究采用

的是自由度更高的界定——只要被试报告不同于实际听觉刺激的感知, 都算是发生了 McGurk 效应(Gurler et al., 2015; Mallick et al., 2015; Wilson, Alsius, Paré, & Munhall, 2016)。这种定义更符合“McGurk 效应反映了视觉信息对听觉感知的影响”这一观点。目前, 多数研究者倾向于采取后者这种高自由度的界定, 以纳入所有视听信息交互的情况(Alsius et al., 2018; Tiippana, 2014)。本文中涉及的研究多数是后一种界定。

3　McGurk 效应的影响因素

3.1　影响 McGurk 效应个体内变异的因素

McGurk 效应的个体内变异是指对同一个体而言, 其 McGurk 效应发生率由于受到某些因素的影响而发生改变的现象(即在被试内设计中, 不同实验条件之间的 McGurk 效应发生率改变)。造成 McGurk 效应个体内变异的因素主要有物理刺激(例如: 视觉、听觉刺激及其同步性等自下而上的外部因素)和认知因素(例如: 注意分配、心理预期等自上而下的内部因素)。

3.1.1　物理刺激因素

视觉刺激变化可能影响个体对视觉信息的加工效果(即影响唇读过程), 进而造成 McGurk 效应的个体内变异。视觉信息呈现的质量越好(越清晰、越完整), 越容易发生 McGurk 效应; 而破坏视觉信息的呈现会降低其对听觉感知的影响, 即减少 McGurk 效应。研究者通过降低视频分辨率(Wilson et al., 2016)、对视频进行马赛克转换(MacDonald, Andersen, & Bachmann, 2000)、对视频进行空间像素化处理(Thomas & Jordan, 2002)等技术手段来降低视频的清晰程度。结果均表明, McGurk 效应随视频清晰度的降低而减少。也有研究将视频切分后只呈现其中一部分(Jordan & Thomas, 2011; Ujiie, Asai, & Wakabayashi, 2015)、或是用光点来呈现说话者的面部运动信息(损失了很多原有面部运动信息) (Jordan, McCotter, & Thomas, 2000), McGurk效应(相比于呈现完整的面部视频)也会减少。还有研究在 10 米或 20 米之外呈现视频(距离越远视频越看不清), McGurk 效应会随距离增加而减少(Jordan & Sergeant, 2000)。此外, 将视频里的面孔倒置(人们对倒置的面孔加工更困难) (Thomas & Jordan, 2002), 或是将正立面孔的嘴部倒置(这种奇怪的面孔也会增加人们对面孔的加工难度), McGurk 效应也会减少(Rosenblum, Yakel, & Green, 2000; Ujiie, Asai, & Wakabayashi, 2018)。近来还有研究发现, 降低视频的播放速度(这可能破坏原本流畅的视觉信息)也会减少 McGurk 效应(Magnotti, Mallick, & Beauchamp, 2018)。

虽然视觉信息的呈现质量对 McGurk 效应影响较大, 但通过破坏视觉信息很难完全消除 McGurk 效应。只要仍有少量有效的视觉信息线索, 效应都还会发生。即 McGurk 效应较为稳定、不易消除。研究表明, 即使呈现马赛克程度最高的视觉信息(MacDonald et al., 2000), 或是将面部距离增加到 20 米之远(Jordan & Sergeant, 2000), 甚至将视频的嘴部区域删除(Jordan & Thomas, 2011), McGurk 效应仍会发生。

不过, 如果视觉信息没有被意识觉察(阈下呈现), 就不会发生 McGurk 效应。即对于 McGurk 效应而言, 阈下视觉加工不足以引发视听整合(Munhall, ten Hove, Brammer, & Paré, 2009), 视觉信息需要被意识觉察才可能引起 McGurk 效应。有研究使用连续闪烁抑制范式(continuous flashing suppression, CFS, Fang & He, 2005; Tsuchiya & Koch, 2005)将 McGurk 刺激的视觉信息呈现在阈下。结果表明, 在 CFS 条件下, McGurk 效应消失了(Palmer & Ramsey, 2012)。还有研究设计了一种动态双歧图的 McGurk 刺激呈现方法(一个花瓶的边缘构成两个面对面的侧脸。花瓶在旋转, 其边缘构成的侧脸在旋转过程中呈现出嘴型的变化。被试对该动态双歧图的感知会在“侧脸”和“花瓶”之间变化)。如果 McGurk 效应的发生无需意识觉察视觉信息, 那么无论个体对双歧图的感知如何, 都应该会发生 McGurk 效应。但如果 McGurk 效应的发生需要意识觉察视觉信息, 则只有在个体对双歧图的感知是“侧脸”时, 才会发生 McGurk 效应(当感知为“花瓶”时, 不会发生 McGurk 效应)。实验结果也支持了后一个推论(Munhall et al., 2009)。

当然, 有的视觉信息对 McGurk 效应影响不大。McGurk 效应只对视觉言语信息(相关面部肌肉的运动)敏感, 只要不影响视觉言语信息的呈现效果, 就不会影响 McGurk 效应。例如: 有研究表明, 无论将视觉刺激用彩色呈现还是用黑白呈现, 其 McGurk 效应发生率都没有差异(Jordan et al., 2000)。

相比于视觉信息, 改变听觉信息影响 McGurk

效应的研究很少。这可能是因为 McGurk 效应本身就是“对听觉信息的感知受到视觉信息的影响而发生变化”, 如果改变听觉刺激, 就难以区分听觉感知发生的变化究竟是来自视觉信息的影响, 还是来自听觉信息本身改变的影响。不过, 仍有研究者从听觉刺激的角度揭示了 McGurk 效应的稳定性——音调、音高等因素对 McGurk 效应的影响不大。他们比较了正常说出音节和唱出音节(用升调、降调两种唱法)对 McGurk 效应的影响, 结果表明在“唱出”和“说出”两种条件下的 McGurk 效应发生率没有显著差异(Quinto, Thompson, Russo, & Trehub, 2010)。

还有研究者针对听觉信息的呈现来拓展 McGurk 效应的研究范式。他们在视觉刺激不变的情况下, 改变听觉刺激的呈现条件。即视觉刺激总是“ba”, 而听觉刺激可能是“ba”(与视觉信息一致), 也可能是一种听起来像“a”的音频(将“ba”的辅音信息减弱)。如此一来, 后者的刺激组合也会诱发被试报告听到了“ba” (但实际的听觉刺激是“a”), 即视觉言语信息对听觉感知形成了“补充”。这与经典 McGurk 效应类似(Irwin, Avery, Brancazio, Turcios, Ryherd, & Landi, 2018)。该范式可以归为 McGurk 范式的一种变式——经典 McGurk 效应关注的是听觉信息不变, 改变视觉信息可能改变个体的听觉感知; 而该变式关注的是视觉信息不变, 改变听觉信息后, 视觉信息会对听觉感知进行补充, 也体现了视觉信息影响听觉感知。未来研究可以尝试将该范式与传统的 McGurk 范式进行比较, 验证二者是否有类似的机制(例如两种范式的效应发生率是否相似? 是否激活了相似的视听整合相关脑区?), 可考虑将该变式作为另一个视听整合的指标。

最后, 视觉和听觉刺激呈现的同步性也可能造成 McGurk 效应的个体内变异。在视听整合研究中, 视觉和听觉刺激不一定要精确地同步呈现才会引起视听整合, 在一定时间窗内的视听刺激异步对视听整合影响不大(Munhall, Gribble, Sacco, & Ward, 1996; Stevenson, Zemtsov, & Wallace, 2012)。McGurk 效应也不例外。研究发现, 只要听觉刺激(相比于视觉刺激)呈现的延迟在–360~360 ms 的时间窗内, 都会产生 McGurk 效应。当然, 同步性的降低同时也会导致 McGurk 效应减少(Munhall et al., 1996)。此外, 即使被试能够感知到视听信息呈现的不同步, 也仍然可能产生 McGurk 效应(Soto-Faraco & Alsius, 2009), 这也体现了 McGurk 效应的稳定性。

总体而言, McGurk 效应一方面容易受到物理刺激因素影响而发生个体内变异, 但另一方面又具有较强的稳定性(不容易完全消失)。现有研究大都关注自下而上的物理刺激因素如何影响 McGurk 效应(尤其关注视觉信息的影响), 也得出了较为一致的结论; 然而却忽视了听觉信息的作用。一个值得探究的问题是: 当听觉信息的可靠性下降时(信噪比降低), McGurk 效应如何变化? 这是实际生活中很常见的视听言语感知情景(例如在嘈杂的环境中与别人聊天)。针对这一问题, 我们预期: 由于听觉信息可靠性降低, 个体对视觉信息的权重增加, 即视觉信息对听觉感知的影响增加, 这可能引发更多 McGurk 效应。

3.1.2 认知因素

如上文所述, 物理刺激的改变对 McGurk 效应的影响较大。但即使面对相同的物理刺激, 个体的认知状态不同, 也可能造成 McGurk 效应发生率改变。而且, 相比于物理刺激这类自下而上的调节因素, 自上而下的认知因素变化在实际生活中更常见(例如我们面对的常常是物理刺激相同的面孔, 但自身的认知状态容易发生改变)。然而这类研究并不多。已有研究主要围绕注意分配进行探讨——当个体分配给 McGurk 任务的注意减少时, McGurk 效应就会减少。研究采用双任务范式, 要求被试在进行视听判断任务(McGurk 任务)的同时进行一项无关的视觉或听觉任务(这降低了被试分配在 McGurk 任务上的注意)。结果表明, McGurk 效应发生率在双任务条件下比单任务条件低(Alsius, Navarra, Campbell, & Soto-Faraco, 2005)。进一步研究还发现, 如果被试同时进行一项触觉任务(不同于视觉、听觉通道的第三个感觉通道), 则 McGurk 效应发生率也会降低(Alsius, Navarra, & Soto-Faraco, 2007)。这提示注意分配对 McGurk 效应的影响并不仅仅局限于视觉或听觉通道, 而是受到一般性的注意分配的影响。另一项采用双任务范式的研究让被试同时进行一项工作记忆任务, 也发现了一致的结果(Buchan & Munhall, 2012)。还有研究在呈现面部视觉信息时, 同时呈现一个分心刺激(一片叶子划过面部)。当要求被试忽略面部去注意分心刺激时(相比于要求

被试忽略分心刺激去注意面部的情况), McGurk 效应的发生率更低(Tiippana, Andersen, & Sams, 2004)。

除了注意分配, 还有研究探讨了预期对 McGurk 效应的影响——如果明确告诉被试接下来呈现视听一致刺激(但实际上仍会包含视听不一致的 McGurk 刺激), 相比于告知被试视听刺激可能不一致的情况, McGurk 效应的发生率更高(Gau & Noppeney, 2016)。即个体预期视听一致会促进 McGurk 效应的发生。

综上所述, 在 McGurk 效应的个体内变异研究中, 研究者更多关注自下而上的物理刺激因素对 McGurk 效应的影响, 但对自上而下的认知相关因素关注较少。虽然已有研究探讨注意分配和预期如何影响 McGurk 效应, 但这一方向仍有较大的发展空间。未来可以考虑探究其它自上而下的认知因素, 例如个体的情绪状态对 McGurk 效应的影响——在不同的情绪状态下, 个体的视听整合或许会发生变化, 这也更贴近日常视听言语感知情景。

另一个生活中常见但却研究较少的问题是: 面孔本身的社会属性如何影响视听言语感知。我们常常与不同的人交流, 而不同人的面孔具有不同的社会属性(面孔情绪、吸引力、重要性、熟悉度等), 这与视觉言语信息加工可能发生交互, 进而影响言语感知。有研究探讨了面孔熟悉度、以及声音面孔是否匹配对 McGurk 效应的影响, 结果表明, 当声音与面孔不匹配时, 对面孔熟悉的被试感知到更少的 McGurk 效应 (Walker, Bruce, & O'Malley, 1995)。另一项研究发现, 如果将不同情绪的声音和面部一起呈现, 要求被试判断声音的情绪, 那么被试的判断会受到面部情绪的影响而产生偏差。而且当对听觉信息的性别进行判断时, 被试也会受到视觉信息性别的影响(de Gelder & Vroomen, 2000)。所以, 我们有理由推测, 在 McGurk 效应中, 即使不改变视觉信息的物理特性, 面孔本身就具有的社会属性也可能影响听觉感知, 这值得进一步研究。最近, 我们尝试探究了与奖赏联结的面孔如何影响 McGurk 效应。结果表明, 相比于没有与奖赏联结的面孔, 与奖赏联结的面孔 McGurk 效应发生率更高。

3.2　影响 McGurk 效应个体间变异的因素

McGurk效应的个体间变异(即个体差异)指的是在同样的测量条件下, 不同个体的 McGurk 效应发生率仍会有差异的现象(即在被试间设计中, 不同组别之间的 McGurk 效应差异)。研究表明, 虽然 McGurk 效应在不同测量条件下可能发生个体内变异, 但如果测试条件相同, McGurk 效应发生率在个体内是较稳定的。对同一批被试间隔 1 年的两次同等条件测量的皮尔逊相关为 0.91 (Mallick et al., 2015); 另一项间隔 2 个月的测量相关为 0.77(Strand, Cooperman, Rowe, & Simenstad, 2014)。但是, McGurk 效应在不同个体间就没那么稳定了。Mallick 等人(2015)测试了 165 名被试, 结果表明不同个体的 McGurk 效应发生率有很大差异(从 0%到 100%)。所以在进行组间比较时, 研究者应谨慎分析组间差异的来源。下文将阐述三个可能与 McGurk 效应个体间变异相关的因素: 对视听信息的依赖程度差异、视听整合能力及其发展差异、语言文化差异。

3.2.1　对视听信息的依赖程度差异

McGurk 效应的个体差异可能来自个体对视觉或听觉信息的依赖程度差异——对视觉信息依赖程度高的个体更容易受到视觉信息的影响, 进而发生更多 McGurk 效应; 而对听觉信息依赖程度高的个体则更不易受到视觉信息影响, McGurk 效应也更少。研究发现, 高水平音乐家(8~13 年专业音乐训练)相比于没有音乐训练的普通人 McGurk 效应发生率更低, 这可能是因为音乐家通过长期训练培养了出色的听觉能力使其更倾向于使用听觉信息(Proverbio, Massetti, Rizzi, & Zani, 2016)。另一项研究表明, 相比于双眼进行 McGurk 任务的被试, 闭上一只眼睛进行任务的被试 McGurk 效应发生率更低(Moro & Steeves, 2018), 这可能是因为视觉通道部分受阻之后, 个体对听觉通道的依赖程度增加。还有研究发现, 在视听言语感知任务中, 老年人更容易受到视觉信息的影响(即老年人的 McGurk 效应发生率比年轻人高), 这可能是因为随着年龄的增长, 老年人的听觉机能退化得比视觉快, 进而对视觉信息的依赖增强(Sekiyama, Soshi, & Sakamoto, 2014)。

对特殊人群(高自闭特质者、听力受损者、视力受损者)的 McGurk 效应研究也支持上述观点(即对视听信息的依赖程度差异可能造成 McGurk 效应的个体间变异)。研究发现, 自闭症谱系障碍 (autism spectrum disorder, ASD)的儿童在面孔记

忆任务上表现更差，在视听言语感知任务中也更少受到视觉信息的影响，即 McGurk 效应发生率比正常儿童低(de Gelder, Vroomen, & van der Heide, 1991)。这可能是因为 ASD 儿童加工面部整体信息的能力较低，无法有效利用视觉信息(即对视觉信息依赖程度低)。也有研究测量了被试的自闭症谱系商数(autism spectrum quotient, AQ)，结果表明，高 AQ 者的 McGurk 效应比低 AQ 者少(Ujiie et al., 2018)，并且 AQ 得分与 McGurk 效应发生率负相关(Ujiie et al., 2015)，即自闭特质越高，McGurk 效应发生率越低，这也与上述 de Gelder 等人(1991)的结论一致。另一方面，听力受损者(有人工耳蜗植入或配备有助听器)和听力正常者一样会发生 McGurk 效应，但是听力受损者对视觉信息的依赖程度更高，McGurk 效应发生率也更高(Rouger, Fraysse, Deguine, & Barone, 2008)。这一结果在听力受损儿童中得到了重复(石涯，王永华，李文靖，2016)。此外，听力受损者的 McGurk 效应会受到手语的影响：如果手语和唇形一致(但与声音不一致)，则他们更容易报告听到视觉信息的音节(手语或唇形)，这提示他们在视听感知中非常依赖视觉信息(Bayard, Colin, & Leybaert, 2014)。最后，视力受损者(从小失去了一只眼睛)的 McGurk 效应发生率低于单眼(或双眼)进行任务的视力正常者(Moro & Steeves, 2018)，这可能是因为视力受损者更倾向于依赖听觉信息。

总体而言，不同人群之间的比较均体现了视听信息依赖程度对 McGurk 效应的影响。然而，组间比较存在的问题是：除了视听信息依赖程度的差异，McGurk 效应还可能受到其它人群间差异的影响。所以，未来研究可考虑直接操纵影响视听信息依赖程度的因素，提供更完善的因果关系证据。例如：可以考虑将 Moro 和 Steeves (2018)的研究修改为组内设计，即比较同一组个体在单眼进行任务和双眼进行任务时的 McGurk 效应发生率。也可以考虑进行纵向追踪研究(例如：比较乐器学习者学习乐器前后的 McGurk 效应差异)。

3.2.2 视听整合能力及其发展差异

个体在分别接收视听信息后对二者的整合(即视听整合)能力的差异也可能与 McGurk 效应的个体差异有关。整合能力较强者可能更容易发生 McGurk 效应。相对的，整合能力较弱者 McGurk 效应更少。研究表明，视听整合时间窗的范围大小存在个体间差异，并在一定程度上反映了视听整合能力(Stevenson et al., 2012)——个体整合时间窗边界越靠右(即在仍能发生整合的情况下，视觉刺激呈现后，听觉刺激呈现得越晚；也即整合时间窗的范围越大)，该个体发生 McGurk 效应的可能性也越大(Stevenson et al., 2012)，即视听整合能力越强的个体，越容易发生 McGurk 效应。

关于 McGurk 效应的发展研究也支持上述观点(即视听整合能力差异是造成 McGurk 效应个体间变异的因素之一)。研究表明，12 岁前儿童的 McGurk 效应发生率比成人低(Hockley & Polka, 1994; McGurk & MacDonald, 1976)，这可能是因为儿童的视听整合能力尚在发展中(较低)，而成人的视听整合能力已经发展成熟(较高)。不过，即使是 4~5 个月大的还未学会说话的婴儿就已经会发生 McGurk 效应(Burnham & Dodd, 2004; Rosenblum et al., 1997)。即婴儿在学会说话前，视听整合能力就已经开始发展，而且大约 12 岁左右就能发展到成人水平。所以儿童与成人的 McGurk 效应差异可能就是来自视听整合能力的差异。

综上所述，个体整合能力越强、发展越完善，McGurk 效应就越强。然而，大部分研究都以 McGurk 效应本身作为视听整合能力的指标，很少有研究利用别的指标测量视听整合能力，并与 McGurk 效应的测量结果相比较。所以，视听整合能力与 McGurk 效应的关系还需要进一步探究。这样一方面有助于确认视听整合能力差异是否确实是 McGurk 效应个体间变异的来源，另一方面有助于确认利用 McGurk 范式探究视听整合的有效性。值得注意的是，最近有研究发现：个体在噪声中利用视觉信息辅助听觉理解句子的能力(也常被视为视听整合能力的指标)与个体的 McGurk 效应发生率没有显著相关(Van Engen, Xie, & Chandrasekaran, 2017)。这进一步警示我们，McGurk 效应发生率与视听整合能力的关系需要更细致的探讨。未来研究应该采用更多指标(例如上文提到的视听整合时间窗大小、对视听刺激的反应时、以及其它视听整合相关任务等)评价视听整合能力，并探究这些指标与 McGurk 效应的关系。

值得一提的是，上述视听整合能力的发展情况在汉语母语儿童中有不一致的结果。研究发现汉语母语的二年级、五年级小学生以及一年级大学生都表现出 McGurk 效应，但这三类人之间的

McGurk 效应发生率没有差异，即没有表现出上述英语母语者的发展趋势(李燕芳, 梅磊磊, 董奇, 2008)。后续研究发现，汉语母语儿童在视听不一致、视听一致、单独听觉条件下，判断声音刺激的正确率没有差异；但是汉语母语大学生在视听不一致条件下正确率低于单独听觉和视听一致条件，即成人更容易受视觉信息影响(李燕芳, 梅磊磊, 董奇, 2009)。这又与英语母语者的研究结果一致。这些研究体现出了语言文化差异与视听整合能力发展的交互。下文将对语言文化差异的影响进行详细阐述。

3.2.3 语言文化差异

McGurk 效应是一种言语感知现象，具有不同文化背景(使用不同母语)的人在 McGurk 效应上可能存在差异，即语言文化差异也是造成 McGurk 效应个体间变异的因素之一。研究发现日语母语者的 McGurk 效应发生率比英语母语者低(Hisanaga, Sekiyama, Igasaki, & Murayama, 2016; Sekiyama & Tohkura, 1993)。这可能是由于日语母语者相比于英语母语者更少受到面部视觉信息的影响。在日本文化中，注视别人面部是不礼貌的，所以日本人在面对面交流中更倾向于使用听觉信息，而不是视觉信息。后续研究还发现汉语母语者的 McGurk 效应发生率也比英语母语者低(Sekiyama, 1997)。

不过，也有研究者没有发现汉语、英语母语者之间的 McGurk 效应差异(Magnotti, Mallick, Feng, Zhou, Zhou, & Beauchamp, 2015)。他们认为 McGurk 效应本身就有较大的个体差异，组间比较的样本不宜太少，于是采用较大样本(307 人)、较多 McGurk 刺激(9 个)进行测量。结果表明 McGurk 效应发生率在汉语、英语母语者人群内部有较大的个体差异，但在两类人群之间整体而言没有显著差异。

除了 McGurk 效应发生率的差异，不同语言文化背景还可能影响个体在发生 McGurk 效应时感知到的音节类型。研究发现，对于经典的 McGurk 刺激(视觉“ga”听觉“ba”)，英语母语者更多报告感知到“tha”，而日语母语者更多报告感知到“da”。这可能与母语差异有关——日语中并没有“th”的发音，而英语日常生活中“tha”的发音多于“da”的发音(Burnham & Dodd, 2018)。

总体而言，语言文化差异影响 McGurk 效应发生率的研究结果不一致。其中获得阳性结果的研究样本量较小，而大样本研究没有发现显著差异。考虑到 McGurk 效应发生率本身具有较大的个体差异，所以语言文化因素究竟是不是 McGurk 效应个体差异的来源，仍旧存疑。一种解释是：语言文化差异确实会对视听言语感知产生影响(例如上文提到的音节感知类型差异)，只是对 McGurk 效应发生率的影响不够明显。这可能是因为不同语言文化背景者对 McGurk 刺激的加工趋于某个相似的“阈限”——有研究表明，即使 McGurk 效应没有发生，视觉信息也已经对听觉感知产生了影响(Brancazio & Miller, 2005)。所以 McGurk 效应的发生可能是连续的过程，视觉信息的影响需要达到一定程度才会产生效应(即存在某个“阈限”)。在世界文化交融的当今社会，各国大学生被试在视听言语感知中对视觉信息的加工越来越相似，即达到 McGurk“阈限”的程度越来越相似，故难以体现出文化差异。所以未来研究除了考虑扩大样本量之外，还应该选取更为典型的语言文化群体(而不是容易接触到不同文化的大学生群体)，或许会有进一步发现。

4 McGurk 效应的认知神经机制

4.1 McGurk 效应的眼动模式

动态人脸是一种包含很多信息的复杂刺激，那么导致 McGurk 效应发生的视觉信息究竟是人脸的什么信息？研究者们尝试采用眼动实验来探究此问题。目前的研究结果提示：引发 McGurk 效应的视觉信息主要来自人脸的嘴部区域。但对嘴部的直接注视不是引起 McGurk 效应的必要条件。除嘴部之外，面部的其它区域同样能提供少量但有效的视觉言语信息，进而引发 McGurk 效应。

在言语感知中，视觉言语信息主要来自嘴部区域的运动。所以引发 McGurk 效应的视觉信息也主要来自嘴部区域。有研究探讨了眼动模式的个体差异与 McGurk 效应个体差异的关系。结果表明，容易产生 McGurk 效应的个体看嘴部区域的时间更长，且看嘴部区域的时间与 McGurk 效应发生率正相关(Gurler et al., 2015)。类似的，英语母语者的 McGurk 效应发生率比日语母语者高，而英语母语者看嘴部区域的时间也更长(Hisanaga et al., 2016)。另一方面，采用双任务范式的研究发现，相比于单任务条件，在双任务条件下 McGurk

效应发生率更低，并且被试对视觉刺激的面部区域注视更少，对嘴部区域的注视也更少(Buchan & Munhall, 2012)。

但是，也有不一致的结果——研究发现，被试是否看嘴部区域与 McGurk 效应的变化并没有关系(Hisanaga et al., 2016; Paré, Richler, ten Hove, & Munhall, 2003; Wilson et al., 2016)。这提示对嘴部区域的中央视野加工对 McGurk 效应的发生并不是必须的，外周视野就能获取足够诱发 McGurk 效应的嘴部视觉言语信息。例如：Paré 等人(2003)进行的一系列实验发现，McGurk 效应的感知与个体注视点是否在嘴部区域没有相关。他们还直接控制了个体的注视点位置，结果表明，只要个体的注视点还在面部区域内，无论是注视嘴部、眼睛、还是额头，都不影响 McGurk 效应发生率。只有当个体注视点离开嘴部区域 10°~20°时，McGurk 效应才会显著减少(但仍然存在)，只有离开嘴部区域 60°以上，McGurk 效应才会完全消失。

考虑到上述不一致的研究结果，嘴部区域注视时间与 McGurk 效应发生率的关系还需要进一步探究。已有研究结果不一致可能有两个原因：(1)不同研究之间使用的研究范式或分析方法不同。例如：Buchan 和 Munhall (2012)比较的是双任务和单任务条件下的人群内差异；Gurler 等人(2015)比较的是自由注视状态下的人群间差异；Paré 等人(2003)的研究不是自由注视(他们尝试控制被试的注视位置)，并且记录眼动的方法与其他研究不同(使用粘附人眼角膜的感应线圈，而非其他研究常用的红外捕捉技术)。以上实验设计或操作上的差异都可能导致研究之间结果不同。(2)不同研究之间的兴趣区划分方法存在差异。例如 Gurler 等人(2015)以及 Buchan 和 Munhall (2012)采用的是方形兴趣区，而 Wilson 等人(2016)则采用圆形兴趣区，这也可能影响注视时间的结果。

除了嘴部区域，面部其它区域同样能提供足以诱发 McGurk 效应的视觉言语信息。研究发现，即使不呈现嘴部区域(将视频沿对角线切分，只呈现没有嘴部的那一部分；或将视频沿水平中轴切分，只呈现上半部分)，McGurk 效应也不会完全消失(Jordan & Thomas, 2011)。在使用其它范式的视听整合研究中也发现了类似的效应——即使消除嘴部运动信息(只留下面部其它区域的运动信息)，视听整合仍然会发生(Thomas & Jordan, 2004)。遗憾的是，这些研究均没有采用眼动技术。而在其它采用眼动技术的 McGurk 效应研究中，研究者都只关注了嘴部以及眼睛区域，忽略了面部其它区域。所以未来研究除了关注嘴部区域，还应该比较面部其它区域的眼动差异(例如鼻子、脸颊等嘴部周边区域。即在保证兴趣区大小基本一致的前提下，尽量让所有兴趣区覆盖整个面部区域)。这可能为我们进一步理解 McGurk 效应提供证据。例如：我们最近的一项研究表明，与奖赏联结的面孔(相比于未与奖赏联结的面孔)发生更多 McGurk 效应，且被试对其嘴部周边区域(鼻子、脸颊)的注视时间更长、注视点个数更多；但对嘴部区域的注视时间却反而更短、注视点个数更少。该结果也支持了上文提到的推论(面部其它区域也能提供有效的视觉言语信息；而对嘴部区域的注视不是发生 McGurk 效应的必要条件)。

4.2 McGurk 效应的加工阶段

大脑接收了视听信息的输入之后，开始对其进行整合加工。此时涉及的问题是：大脑在接收刺激后的不同阶段里如何加工视听刺激，进而产生 McGurk 效应？研究者们尝试用具有较高时间分辨率的脑电技术(electroencephalogram, EEG)或是脑磁图技术(magnetoencephalography, MEG)回答该问题。目前的研究结果提示：对视听信息的整合发生在加工早期阶段；而在加工晚期阶段，大脑会尝试解决 McGurk 刺激的视听不一致冲突。

发生 McGurk 效应时，视听整合过程在加工早期就已经发生。研究发现，对于 McGurk 刺激而言，当发生 McGurk 效应时，N1 波幅相比于视听一致刺激更小；而且相比于没有发生 McGurk 效应的 McGurk 刺激也更小(Romero, Senkowski, & Keil, 2015)。N1 主要由听觉刺激造成。相比于单独听觉刺激，视听刺激引发的 N1 波幅更小，这可能反映了视听整合过程中视觉信息利用率的增加(Besle, Fort, Delpuech, & Giard, 2004)。所以，发生 McGurk 效应时的 N1 波幅降低可能提示了此时视觉信息对听觉信息的影响更明显。而且 N1 是事件相关电位(event related potential, ERP)的第一个负波，这也提示这种影响发生在加工早期阶段。神经振荡结果也表明，当 McGurk 效应发生时，Beta 频段的抑制相比于视听一致的刺激在加工早期(0~500 ms)更强 (Romero et al., 2015)。这与上述 N1 结果类似，提示了 McGurk 效应的发生(相

比于视听一致的情况)需要更强的视听整合, 而且这种整合在加工早期就已经发生。

采用 oddball 范式进行的研究也支持 McGurk 效应中的视听整合过程发生在加工早期的观点。这类研究将视听一致刺激作为标准刺激, McGurk 刺激作为偏差刺激, 比较 McGurk 刺激和视听一致刺激的 ERP。结果表明, 在早期加工阶段(听觉刺激呈现后 200~300 ms), McGurk 刺激会诱发失匹配负波(mismatch negativity, MMN) (Saint-Amour, De Sanctis, Molholma, Ritter, & Foxe, 2007)。MMN 反映了对出现频率较低的新异听觉刺激的探知; 反映了大脑将当前听觉刺激与之前的一系列听觉刺激进行比较的加工过程。当听觉感知改变时, 就会产生 MMN。因此, MMN 可以作为听觉辨别能力的电生理指标。MMN 常由听觉刺激的物理属性改变而诱发。不过, 对 McGurk 刺激而言, 听觉刺激的物理属性没有发生变化, 但主观感知变化也引起了 MMN, 这被称为 McGurk-MMN。McGurk-MMN 在很多研究中得到了重复(Colin et al., 2002; Colin, Radeau, Soquet, & Deltenre, 2004; Eskelund, MacDonald, & Andersen, 2015)。这提示在加工 McGurk 刺激的早期阶段, 个体就已经感知到了新异的听觉刺激(虽然听觉刺激的物理属性实际上并没有改变), 即视听整合已经发生了(辛昕, 任桂琴, 李金彩, 唐晓雨, 2017)。采用类似 oddball 范式的 MEG 研究也表明, 发生 McGurk 效应时, 在早期加工阶段(听觉刺激呈现后 160 ms 以及 270 ms), 不同脑区的 Gamma 振荡活动会增强(Kaiser, Hertrich, Ackermann, Mathiak, & Lutzenberger, 2005)。而 Gamma 神经振荡与信息整合加工有关(钱浩悦, 黄逸慧, 高湘萍, 2018)。这与上述 McGurk-MMN 的结果类似, 即此时视听整合已经开始发生, 大脑探测到了听觉信号的“改变” (实际上是主观感受改变, 物理刺激并没有变化)。

在上述 McGurk-MMN 的相关研究中(采用 oddball 范式), 研究者主要比较了 McGurk 刺激(偏差刺激)与视听一致刺激(标准刺激)的差异, 但是这类研究忽略了以下问题: 此时产生的 MMN 究竟是由于听觉感知发生改变, 还是由于视觉信息与听觉信息不一致? 所以, 在未来的相关研究中应该考虑加入一组对照条件——将视听一致刺激作为标准刺激、视听不一致且不会诱发 McGurk 效应的刺激作为偏差刺激。依据听觉感知发生改变才会发生 MMN 的假设, 这种情况下的视听不一致刺激相比于视听一致刺激应该不会产生 MMN。这一推论值得进一步探究。

在加工的相对晚期阶段(上述视听整合过程已经开始之后), 大脑会尝试解决视听不一致冲突(McGurk 刺激的视听信息实际上是不一致的, 所以可能发生冲突)。研究表明, 在刺激呈现后 500~800 ms, McGurk 刺激相比于视听一致刺激有更强的 Beta 频段抑制。依据已有研究, 视听不一致刺激的 Beta 频段抑制比视听一致刺激强(Lange, Christian, & Schnitzler, 2013), 这可能反映了视听不一致的冲突效应以及自上而下的冲突解决过程。即大脑可能在加工的相对晚期阶段才探测到视听不一致冲突, 并且尝试解决。另一方面, 采用 oddball 范式的 MEG 研究也表明, 发生 McGurk 效应时, 加工晚期的 Gamma 频段活动会增强, 这也提示了与听觉信息不一致的视觉信息对听觉感知的影响(Kaiser et al., 2005)。有趣的是, 即使被试报告感知到视听不一致, McGurk 效应仍会发生(Soto-Faraco & Alsius, 2009)。这提示, 即使视听不一致冲突没有解决, 视听整合也会发生, 二者是相对独立的过程。

4.3　McGurk 效应的相关脑区

除了时间进程问题, 在大脑加工 McGurk 刺激的过程中, 另一个重要问题是: 哪些脑区参与了加工以及这些脑区起何作用? 研究者们尝试用具有较高空间分辨率的功能性磁共振成像技术(functional magnetic resonance imaging, fMRI)、经颅磁刺激技术(transcranial magnetic stimulation, TMS)和 MEG 回答此问题。目前的研究结果提示: 颞上皮层(superior temporal cortex)与视听整合过程相关; 额下皮层(inferior frontal cortex)与视听不一致冲突相关。

在发生 McGurk 效应的过程中, 颞上皮层与视听整合密切相关(Beauchamp et al., 2010; Miller & D'Esposito, 2005; Nath & Beauchamp, 2012)。早期 fMRI 研究表明, 相比于没有发生 McGurk 效应, 当发生 McGurk 效应时, 颞上皮层的激活更强(Jones & Callan, 2003)。对 McGurk 效应个体差异的神经基础研究发现, McGurk 效应发生率在 50% 以上的被试(强 McGurk 感知者)相比于发生率在 50%以下的被试(弱 McGurk 感知者), 左侧颞上沟

(left superior temporal sulcus, lSTS)的激活更强，且其激活程度与 McGurk 效应发生率有显著正相关(Nath & Beauchamp, 2012)。该结果在 6~12 岁儿童的研究中得到了重复(Nath, Fava, & Beauchamp, 2011)。更重要的是，Beauchamp 等人(2010)使用 fMRI 技术定位每个被试的 STS，之后使用 TMS 抑制 STS 的激活。结果表明，使用 TMS 刺激 STS 之后，被试的McGurk效应发生率降低了，但是对一般视听材料的判断不受影响。类似的，Marques, Lapenta, Merabet, Bolognini 和 Boggio (2014)使用经颅电刺激技术(transcranial direct current stimulation)刺激 STS，也得到了与 Beauchamp 等人(2010)一致的结果。在 EEG 研究中，Saint-Amour 等人(2007)对上文提到的 McGurk-MMN 进行了溯源分析，发现了左侧颞叶皮层的主导效应。MEG 研究也发现，在发生 McGurk 效应之前会伴随着多个脑区的神经振荡，尤其是左侧颞上回(left superior temporal gyrus)的 Beta 神经振荡，研究者认为这提示了视听整合的过程(Keil, Müller, Ihssen, & Weisz, 2012)。

McGurk 效应与颞上皮层的关系研究结果较为一致，但仍有进一步探索的空间。最近，一项视听整合的研究发现，STS 对视听整合的反应可以再细分：STS 的某些体素(voxels)对面孔的嘴部运动更敏感，而另一些体素对面孔的眼部运动更敏感。当视听信息呈现时，STS 激活，且只有对嘴部运动敏感的体素会对听觉刺激有较强的反应。这提示 STS 脑区在整合视听信息的过程中，视觉和听觉信息都会一起加工，但是对整合影响较大的视觉信息(例如嘴部运动)相比于对整合影响较小的视觉信息(例如眼部运动)在其中的加工方式可能不同(Zhu & Beauchamp, 2017)。该研究提示，对 McGurk 效应而言，STS 的激活也可能有类似的效应(例如：对嘴部运动敏感的体素或许可以预测 McGurk 效应发生与否，而对眼部运动敏感的体素则不能)。未来值得从细分脑区激活模式的角度进一步探讨 STS 在 McGurk 效应中的作用。

除了颞上皮层，另一个备受关注的 McGurk 效应相关脑区是额下皮层。该脑区与视听不一致冲突有关(Fernández et al., 2017; Gau & Noppeney, 2016; Nath & Beauchamp, 2012)。在早期的 McGurk 效应 fMRI 研究中就发现了额下皮层的激活(Jones & Callan, 2003)。在 MEG 研究中也发现了左侧额下皮层的神经振荡活动增强(Kaiser et al., 2005)。对 McGurk 效应个体差异的神经基础研究也发现，相比于视听一致刺激，额下回(inferior frontal gyrus, IFG)对视听不一致刺激(包括 McGurk 刺激)的激活更强。但是 IFG 的激活在强 McGurk 感知者和弱 McGurk 感知者之间没有差异。研究者由此推断：IFG 可能与视听不一致冲突有关，但与视听整合过程关系不大(Nath & Beauchamp, 2012)。还有研究发现，相比于没有发生 McGurk 效应的情况，当发生 McGurk 效应时, IFG 的激活更强。而且与冲突探测相关的脑区——前扣带回(anterior cingulate cortex, ACC)的激活也更强(Fernández et al., 2017)。这也提示了 McGurk 效应中存在视听不一致冲突的过程。

Gau 和 Noppeney (2016)的研究也涉及额下皮层激活模式与 McGurk 效应的关系，但与上述 Fernández 等人(2017)的研究结果不一致。具体而言，Gau 和 Noppeney (2016) 使用 fMRI 探究预期对 McGurk 效应的影响。在该研究中，研究者明确告诉被试这一组刺激的视听信息是一致还是不一致(即“告知一致”和“告知不一致”条件)。结果表明，相比于告知不一致条件，在告知一致条件下，被试的 McGurk 效应发生率更高(即被试预期刺激是视听一致时更容易发生 McGurk 效应)。在神经层面，左额下沟(left inferior frontal sulcus, lIFS)在视听不一致时(相比于视听一致)激活更强，这与上述 Fernández 等人(2017)的结果相似。但当被试发生了 McGurk 效应(相比于没有发生 McGurk 效应)时，lIFS 激活减弱。而且，这种效应在被试预期视听一致(发生更多 McGurk 效应)时比预期视听不一致(发生更少 McGurk 效应)时更明显。这似乎与 Fernández 等人(2017)的结果相反——Fernández 等人(2017)发现：发生 McGurk 效应时，IFG 激活更强。

即使有不一致的研究结果，仍可以肯定的是：额下皮层在 McGurk 效应中与视听不一致冲突有关。只是目前还需要进一步探究其激活模式。上述研究结果不一致可能有三个原因：(1)两项研究的范式不同。Fernández 等人(2017)关注的是自然状态下的 McGurk 刺激感知；而 Gau 和 Noppeney (2016)关注的是有心理预期条件下对 McGurk 刺激的感知。即后者可能还包括了预期的效应。(2) fMRI 无法细致区分加工的时间进程。额下皮层确实与视听不一致冲突有关，但是其在冲突解决的

过程中可能有不同的激活模式。具体而言：大脑探测到冲突并刚开始尝试解决时，额下皮层激活增强；而激活越强，就越有利于冲突解决，进而有利于 McGurk 效应的发生。此时比较 McGurk 效应发生和没发生时的额下皮层激活程度，就可能得到 Fernández 等人(2017)的结果。但当过了大脑尝试解决冲突的时间段，如果发生了 McGurk 效应，则可能冲突已经基本解决。所以由于冲突变弱，额下皮层的激活也就随之减小。相对的，如果没有发生McGurk效应，则冲突还没有解决，其激活可能仍然较强。此时比较 McGurk 效应发生和没发生时的额下皮层激活程度，就可能得到 Gau 和 Noppeney (2016)的结果。(3)额下皮层的不同区域可能在不同的时间进程上起到不同的作用。Fernández 等人(2017)定位的是 IFG，而 Gau 和 Noppeney (2016)定位的是 IFS，位置稍有区别。二者可能在上述加工时间进程中起到承接的作用——随着冲突解决程度的改变，额下回的激活模式也随之改变。这个问题值得进一步采用时间、空间分辨率都较高的 MEG 技术深入探究。

综上所述，对 McGurk 效应的相关脑区分析仍有较大探索空间。除了上述额下皮层激活模式之外，未来研究还可以考虑进行功能连接分析。例如对刺激的加工是如何在颞上皮层与额下皮层二者之间传递的？这有助于我们理解 McGurk 效应中的视听整合过程和视听不一致冲突过程。还可以考虑进行多体素模式分析(multivoxel pattern analysis，MVPA)，以探究 McGurk 刺激相比于视听一致刺激或是不能诱发 McGurk 效应的视听不一致刺激的大脑激活模式有何差异。这有助于我们进一步理解大脑对 McGurk 刺激的加工相比于其他视听刺激有何本质差别。

5 总结与展望

McGurk 效应反映了视觉信息对听觉感知的影响。该效应提出至今 40 多年，仍旧是视听言语感知研究中的热点问题。本文尝试对 McGurk 效应的研究要点进行系统性梳理，概括如下：(1) McGurk 效应的测量与界定：诱发 McGurk 效应需要特定辅音的视频和特定辅音的音频组合。目前较常用视觉辅音“g”和听觉辅音“b”的组合。相关研究中最普遍的因变量指标为 McGurk 效应发生率，即对 McGurk 刺激实施多次测量后计算其中发生 McGurk 效应的次数比例。多数研究将 McGurk 效应界定为：只要感知到不同于实际听觉刺激的音节，就算是发生了 McGurk 效应。(2) McGurk 效应的影响因素：包括物理刺激(例如：视觉、听觉刺激、视听刺激异步性)、认知因素(例如：注意分配、心理预期)等造成个体内变异的因素。还包括视听信息依赖程度、视听整合能力、语言文化差异等造成个体间变异的因素。(3) McGurk 效应的认知神经机制：McGurk 效应发生时，视觉言语信息主要来自说话者的嘴部区域(不过，说话者面部其它区域也能提供有效的视觉言语信息)。视听整合过程发生在加工早期阶段、与颞上皮层有关。视听不一致冲突发生在加工晚期阶段、与额下皮层有关。

虽然前人研究对 McGurk 效应进行了细致深入的探讨，但仍然存在一些问题与不足，这在上文已经有所讨论(例如：现有研究很少关注面孔社会属性对 McGurk 效应的影响，也很少关注面部其它区域提供的视觉言语信息，而且眼动和 fMRI 研究中存在不一致的结果等)。下文将从 McGurk 效应中单通道信息加工与视听整合的关系、McGurk 效应的刺激间变异、与计算模型的关系、对后续认知过程的影响、以及范式的标准化与推广性出发，结合已有研究的不足，提出未来研究的可能方向。

5.1 McGurk 效应中单通道信息加工与视听整合的关系

视听整合过程应该涉及两个方面：一是加工外界输入的单通道的视觉和听觉信息；二是对输入的视听信息进行整合。遗憾的是，很少有研究细致区分 McGurk 效应发生率的改变究竟是来自哪个方面，大部分研究只是粗略地解释为“某因素影响了视听整合过程”，而没有进一步讨论该因素究竟是直接影响了视听整合能力本身，还是影响了个体对单通道信息的加工过程(视听整合能力可能不变)，进而影响了视听整合的程度。这是未来研究在解释 McGurk 效应发生率的变化时需要注意的问题。换言之，虽然研究者们公认发生 McGurk 效应就是发生了视听整合，但是直接把 McGurk 效应发生率等同于视听整合能力显得过于武断。因为 McGurk 效应发生率(即视听整合的程度)除了与个体视听整合能力有关之外，还与个体对单通道信息(视觉、听觉信息)的加工有关

(也见本文 3.2)。相应的, 在神经机制方面, 已有研究大都关注 McGurk 效应中的视听整合过程(最近也有研究开始关注 McGurk 效应与视听不一致冲突, 见本文 4.3), 但很少有研究关注对单通道信息的加工在 McGurk 效应神经机制中所起的作用, 这在未来同样值得进一步探讨。

以对视觉信息的加工过程(即唇读过程)为例——我们推测, McGurk 效应的发生与否可能与个体对视觉信息的加工策略(倾向于自上而下地控制还是自下而上地反应)有关, 这一假设主要基于唇读的神经机制研究。研究表明, McGurk 效应的发生率与唇读能力显著正相关(Strand et al., 2014)。而听力正常者唇读过程的神经机制与视听整合过程很相似——唇读与颞上皮层的激活相关(Macsweeney, et al., 2000)。然而, 听力障碍者的唇读却是与海马和后部扣带皮层的活动相关, 而非颞上皮层(Macsweeney, et al., 2002)。其中, 海马的激活提示了记忆在唇读中的重要作用, 而后部扣带皮层则可能是负责将记忆中的语言知识与外部输入的视觉信息进行比较, 进而完成言语感知。这提示听力障碍者在对视觉信息的加工过程(即唇读过程)中更倾向于采取自上而下的加工策略。而听力正常者可能只在更困难的言语加工情境下(例如有噪音时)才调动这种自上而下的加工(张明, 陈骐, 2003)。所以, 我们推测, 不同加工策略并不是非此即彼, 而是连续变化、有所权重, 而个体加工视觉信息时采取的两种加工策略的权重可能与 McGurk 效应有关。

5.2 McGurk 效应的刺激间变异

McGurk 效应存在较大的刺激间变异。即不同的 McGurk 刺激(例如不同的说话人、不同的视听音节组合)对同一个被试而言, 其 McGurk 效应发生率可能有较大差异(Mallick et al., 2015)。目前大部分研究都只采用 1 个或 2 个 McGurk 刺激, 所以在进行研究之间的比较时, 刺激间的变异也可能导致研究结果差异。但很少有研究者考虑这个问题。未来研究可以考虑使用多个 McGurk 刺激, 以期降低 McGurk 效应的刺激间变异的影响。不过, 这样也会带来另一个问题: 如何控制本研究中的刺激间变异。

研究者可以考虑使用 McGurk 效应的差异噪声编码模型(noisy encoding of disparity model, NED, Magnotti & Beauchamp, 2015)来分离 McGurk 效应的刺激间变异。该模型认为不同个体受视觉信息影响的程度、以及表征视听信息的清晰度不同, 不同刺激引起 McGurk 效应的“能力”也不同(有的刺激更容易诱发 McGurk 效应, 有的更不容易), 这些因素共同影响 McGurk 效应是否发生。相应地, NED 模型包括三个参数: 感知噪声(sensory noise, σ)、区别阈限(disparity threshold, T)、刺激差异(stimulus disparity, D)。其中, 感知噪声(σ)描述了个体在表征视听信息时的清晰、准确程度。感知噪声越低, 表征越清晰。区别阈限(T)描述了个体依据视觉信息进行判断的倾向高低。区别阈限越高, 个体越倾向于依赖视觉信息进行判断(即更可能产生 McGurk 效应)。感知噪声和区别阈限都是描述个体间变异的参数。而刺激差异(D)描述了单个 McGurk 刺激引起 McGurk 效应的可能性大小, 是描述刺激间变异的参数。该模型区分了刺激引起的变异和个体的内部差异, 这让研究者可以利用该模型分离出由刺激的差异带来的 McGurk 效应变异。所以, 未来研究可以考虑采用多个 McGurk 刺激、并使用 NED 模型来控制刺激间差异的影响。可以考虑在经过预实验之后, 筛选出刺激差异相似的 McGurk 刺激。也可考虑不直接比较 McGurk 效应发生率, 而是比较模型拟合后的个体相关参数, 即感知噪声和区别阈限的变化。这样一方面可以增加结论的可推广性, 另一方面可以控制由于增加 McGurk 刺激数量而带来的刺激差异混淆。尤其是涉及使用不同刺激进行组间比较的实验、或是不同刺激在被试间交叉平衡的实验。

5.3 McGurk 效应的脑机制与计算模型

除了上文探讨的脑机制相关研究, 计算模型研究也尝试从新的角度对 McGurk 效应的机制进行解释(Marques et al., 2016; Samuel, 2011)。例如上文已经提到的分层预测编码模型(Olasagasti et al., 2015)以及 NED 模型(Magnotti & Beauchamp, 2015)。未来研究应考虑将脑科学技术与计算模型相结合。不同于通过实验操纵或是利用神经生理技术来探究机制的方法, 计算模型研究尝试先假定其中的加工过程, 并利用不同的参数来描述不同的加工过程, 参数在其中代表的意义与特定加工过程相对应。这可能为我们理解某个认知过程提供新的思路。但是, 计算模型比较依赖事先对模型的假设, 其参数拟合大多是依据行为结果(例

如 McGurk 效应发生率)或是视听刺激的物理参数，这与其它探讨 McGurk 效应脑机制的研究(例如 EEG、fMRI 结果)关联较小。遗憾的是，很少有 McGurk 效应的研究将神经生理技术与计算模型相结合。所以，未来的计算模型研究可以考虑利用 EEG、fMRI 结果等神经科学指标进行参数拟合，抑或是神经科学研究可以考虑利用计算模型寻找相应参数的对应脑区，为模型的参数找到神经基础。例如：将神经生理结果与 NED 模型相结合，尝试寻找刺激差异(D)、感知噪声(σ)、区别阈限(T)的相关脑区。这有助于我们定位哪些脑区负责编码刺激差异、哪些脑区负责表征视听信息的清晰度、以及哪些脑区负责对视觉信息的利用等。再如：最近提出的多感觉语言感知的因果推断模型(model of causal inference in multisensory speech perception, Magnotti & Beauchamp, 2017)认为，人们在面对多通道信息时并不是直接进行整合，而是先判断这些不同通道的信息是否同源的可能性(因果推断)，并据此给“整合”或“不整合”分配权重——即在面对视听不一致的 McGurk 刺激时，大脑会先判断视听信息是来自同一个人的可能性(以及不是来自同一个人的可能性)，并据此给“整合”或“不整合”命令分配权重、并平均表征。在完成因果推断之后，如果执行“整合”命令，则会产生 McGurk 效应；反之，则不会发生 McGurk 效应。这提示我们，除了视听整合过程和视听不一致冲突之外，在那之前的因果推断过程可能也是发生 McGurk 效应时的一个步骤。为之寻找相关神经基础有助于我们补充、完善对 McGurk 效应机制的理解。

5.4　McGurk 效应对后续认知过程的影响

多数研究都在关注影响 McGurk 效应的因素，或是直接探讨 McGurk 效应的机制，很少有研究关注 McGurk 效应发生之后的“后续影响”。即 McGurk 效应是否以及如何影响其他认知过程。围绕这个要点，可以提出很多有趣的研究问题。例如：有研究发现，当被试感受过 McGurk 刺激之后，在接下来的单独声音判断任务中，被试会更倾向于认为听到的声音是之前看到的嘴型的声音。即 McGurk 效应会重新校准个体对听觉语音的识别(Bertelson et al., 2003)。类似的，另一项研究也发现，当 McGurk 效应发生(听觉“aba”和视觉“aga”被感知为“ada”)之后，对纯听觉“aba”的判断更容易被错误地知觉为“ada” (McGurk 知觉)。而且当这种情况发生时，大脑的听觉皮层的激活模式与实际听到“ada”时更相似(相比于没有把纯听觉“aba”错误地知觉为“ada”的情况)。这提示当感知到 McGurk 效应时，大脑的神经表征模式会从表征“aba”向“ada”转换，这会影响到后续的纯听觉任务(Lüttke, Ekman, van Gerven, & de Lange, 2016)。这些研究都提示，McGurk 效应的发生确实会对后续认知过程产生影响，探究该问题有助于我们更加全面地认识 McGurk 效应。与此相关的另一个有趣问题是：在 McGurk 效应研究中，刺激材料多采用的是无意义音节(例如听觉“ba”和视觉“ga”感知到“da”)。但有少部分研究采用的是词汇刺激(例如：听觉“bait”和视觉“gate”感知到“date”, Alsius et al., 2005, 2007)。那么当采用词汇刺激时，McGurk 效应发生(或没发生)后的语义激活情况如何变化？是激活了听觉词的语义、还是视觉词的语义、还是整合后感知的语义？抑或是所有语义都有激活，只是激活程度不同？这有助于我们理解 McGurk 效应发生后，原本的听觉与视觉刺激在加工过程中如何变化。

5.5　McGurk 效应的范式标准化和推广性

虽然对 McGurk 效应的研究很多，但不同研究之间在细节上存在较大差异，研究范式的标准化是未来需要重视的问题，主要包括：采用标准化刺激、使用一致的 McGurk 效应界定标准、在实验中加入填充试次、报告完整的描述统计结果。Alsius 等人(2018)尝试对 McGurk 效应的强度进行元分析。但在初步筛出的 276 项研究中，最终符合元分析标准的只有 21 项。而在这 21 项研究之中，只有 2 项研究用表格报告了均值标准差；不同研究之间范式的使用也千差万别。而且，考虑到 McGurk 效应的刺激间变异和个体间变异，在确定造成这些变异的主要原因之前(即可能的调节变量)，对 McGurk 效应的强度进行元分析似乎是不可能的。这强烈提示我们：在未来的研究中，应注意以下问题：(1)采用标准化刺激。研究者们应该建立标准 McGurk 刺激的开放数据库，一方面免去自行录制视频的投入，另一方面可以更好地控制 McGurk 效应的刺激间变异，有助于进行研究间的比较。(2)使用一致的 McGurk 效应界定标准。建议采用宽松的 McGurk 效应界定标准。即只要听觉感知不同于实际的听觉刺激，就算是

发生了 McGurk 效应(Alsius et al., 2018; Tiippana, 2014)。(3)在实验中加入填充试次。建议除了视听一致刺激之外，增加单独听觉的条件作为填充试次，以确认在 McGurk 效应中确实是视觉信息对听觉感知造成了影响，而不是被试听觉感知本身的问题(Alsius et al., 2018)。(4)应该报告完整的描述性统计结果，这是将来进行元分析的必要数据。

最后，研究者还需要注意 McGurk 效应的推广性问题——将 McGurk 效应的研究结论推广到视听一致的言语感知情景中时，需要谨慎(Alsius et al., 2018)。因为 McGurk 效应的加工过程无论在现象上还是神经上都与视听一致时的加工过程不完全一样。主要体现在以下研究中：(1)个体对视听一致刺激的加工不涉及视听冲突，但对 McGurk 刺激的加工可能涉及视听不一致冲突的探测和解决(Fernández et al., 2017)。而且 McGurk 效应的发生率与探测视听不一致的能力(分辨真实的视听一致刺激和 McGurk 刺激)有显著负相关(Strand et al., 2014)。(2)相比于 McGurk 刺激，颞上皮层对视听一致刺激更偏好，即对视听一致刺激的激活更强(Lüttke, Ekman, van Gerven, & de Lange, 2015)。(3)个体的 McGurk 效应发生率与个体在噪声中利用视觉信息辅助听觉理解句子的能力没有显著相关。而后者的刺激主要是视听一致刺激。这提示我们 McGurk 效应不一定能直接替代对视听一致刺激的研究(Van Engen et al., 2017)。

参考文献

雷江华，方俊明. (2005). 聋人唇读的大脑机制研究. *心理科学, 28*(1), 10–12.

李燕芳，梅磊磊，董奇. (2008). 汉语母语者视听双通道言语知觉的特点及发展研究. *心理发展与教育*, *24*(3), 43–47.

李燕芳，梅磊磊，董奇. (2009). 视觉言语在汉语母语儿童和成人英语语音知觉中的作用. *心理科学*, *32*(5), 1038–1041.

朴永馨. (2006). *特殊教育辞典 (第二版)*. 北京：华夏出版社.

钱浩悦，黄逸慧，高湘萍. (2018). Gamma 神经振荡和信息整合加工. *心理科学进展, 26*(3), 433–441.

石涯，王永华，李文靖. (2016). 唇读对听障儿童语音识别的帮助作用. *听力学及言语疾病杂志*, *24*(5), 482–485.

文小辉，李国强，刘强. (2011). 视听整合加工及其神经机制. *心理科学进展, 19*(7), 976–982.

文小辉，刘强，孙弘进，张庆林，尹秦清，郝明洁，牟海蓉. (2009). 多感官线索整合的理论模型. *心理科学进展*, *17*(4), 659–666.

辛听，任桂琴，李金彩，唐晓雨. (2017). 早期视听整合加工——来自 MMN 的证据. *心理科学进展*, *25*(5), 757–768.

徐诚. (2013). 唇读研究回顾：从聋人到正常人. *华东师范大学学报(教育科学版), 31*(1), 56–61.

张明，陈骐. (2003). 听觉障碍人群的言语机制. *心理科学进展, 11*(5), 486–493.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Current Biology*, *15*(9), 839–843.

Alsius, A., Navarra, J., & Soto-Faraco, S. (2007). Attention to touch weakens audiovisual speech integration. *Experimental Brain Research*, *183*(3), 399–404.

Alsius, A., Paré, M., & Munhall, K. G. (2018). Forty years after hearing lips and seeing voices: The McGurk effect revisited. *Multisensory Research*, *31*(1-2), 111–144.

Bayard, C., Colin, C., & Leybaert, J. (2014). How is the McGurk effect modulated by cued speech in deaf and hearing adults? *Frontiers in Psychology*, *5*, 416.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *The Journal of Neuroscience*, *30*(7), 2414–2417.

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk after effect. *Psychological Science*, *14*(6), 592–597.

Besle, J., Fort, A., Delpuech, C., & Giard, M. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. *European Journal of Neuroscience*, *20*(8), 2225–2234.

Brancazio, L., & Miller, J. L. (2005). Use of visual information in speech perception: Evidence for a visual rate effect both with and without a McGurk effect. *Perception & Psychophysics*, *67*(5), 759–769.

Buchan, J. N., & Munhall, K. G. (2012). The effect of a concurrent working memory task and temporal offsets on the integration of auditory and visual speech information. *Seeing and Perceiving*, *25*(1), 87–106.

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*(4), 204–220.

Burnham, D., & Dodd, B. (2018). Language-general auditory-visual speech perception: Thai-English and Japanese-English McGurk effects. *Multisensory Research*, *31*(1-2), 79–110.

Colin, C., Radeau, M., Soquet, A., & Deltenre, P. (2004).

Generalization of the generation of an MMN by illusory McGurk percepts: Voiceless consonants. *Clinical Neurophysiology*, *115*(9), 1989–2000.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology*, *113*(4), 495–506.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion*, *14*(3), 289–311.

de Gelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology*, *3*(1), 69–86.

Eskelund, K., MacDonald, E. N., & Andersen, T. S. (2015). Face configuration affects speech perception: Evidence from a McGurk mismatch negativity study. *Neuropsychologia*, *66*, 48–54.

Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature Neuroscience*, *8*(10), 1380–1385.

Fernández, L. M., Macaluso, E., & Soto-Faraco, S. (2017). Audiovisual integration as conflict resolution: The conflict of the McGurk illusion. *Human Brain Mapping*, *38*(11), 5691–5705.

Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *Neuroimage*, *124*, 876–886.

Gurler, D., Doyle, N., Walker, E., Magnotti, J., & Beauchamp, M. (2015). A link between individual differences in multisensory speech perception and eye movements. *Attention, Perception, & Psychophysics*, *77*(4), 1333–1341.

Hisanaga, S., Sekiyama, K., Igasaki, T., & Murayama, N. (2016). Language/culture modulates brain and gaze processes in audiovisual speech perception. *Scientific Reports*, *6*, 35265.

Hockley, N. S., & Polka, L. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. *The Journal of the Acoustical Society of America*, *96*(5), 3309–3318.

Irwin, J., Avery, T., Brancazio, L., Turcios, J., Ryherd, K., & Landi, N. (2018). Electrophysiological indices of audiovisual speech perception: Beyond the McGurk effect and speech in noise. *Multisensory Research*, *31*(1-2), 39–56.

Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *NeuroReport*, *14*(8), 1129–1133.

Jordan, T. R., McCotter, M. V., & Thomas, S. M. (2000). Visual and audiovisual speech perception with color and gray-scale facial images. *Perception & Psychophysics*, *62*(7), 1394–1404.

Jordan, T. R., & Sergeant, P. (2000). Effects of distance on visual and audiovisual speech recognition. *Language and Speech*, *43*(1), 107–124.

Jordan, T. R., & Thomas, S. M. (2011). When half a face is as good as a whole: Effects of simple substantial occlusion on visual and audiovisual speech perception. *Attention, Perception, & Psychophysic*, *73*(7), 2270–2285.

Kaiser, J., Hertrich, I., Ackermann, H., Mathiak, K., & Lutzenberger, W. (2005). Hearing lips: Gamma-band activity during audiovisual speech perception. *Cerebral Cortex*, *15*(5), 646–653.

Keil, J., Müller, N., Ihssen, N., & Weisz, N. (2012). On the variability of the McGurk effect: Audiovisual integration depends on prestimulus brain states. *Cerebral Cortex*, *22*(1), 221–231.

Lange, J., Christian, N., & Schnitzler, A. (2013). Audio-visual congruency alters power and coherence of oscillatory activity within and between cortical areas. *Neuroimage*, *79*, 111–120.

Lüttke, C. S., Ekman, M., van Gerven, M. A., & de Lange, F. P. (2015). Preference for audiovisual speech congruency in superior temporal cortex. *Journal of Cognitive Neuroscience*, *28*(1), 1–7.

Lüttke, C. S., Ekman, M., van Gerven, M. A. J., & de Lange, F. P. (2016). McGurk illusion recalibrates subsequent auditory perception. *Scientific Reports*, *6*, 32891.

MacDonald, J. (2018). Hearing lips and seeing voices: The origins and development of the 'McGurk effect' and reflections on audio–visual speech perception over the last 40 years. *Multisensory Research*, *31*(1-2), 7–18.

MacDonald, J., Andersen, S., & Bachmann, T. (2000). Hearing by eye: How much spatial degradation can be tolerated? *Perception*, *29*(10), 1155–1168.

Macsweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., ... Brammer, M. J. (2000). Silent speechreading in the absence of scanner noise: An event-related fMRI study. *Neuroreport, 11*(8), 1729–1733.

Macsweeney, M., Calvert, G. A., Campbell, R., McGuire, P. K., David, A. S., Williams, S. C. R., ... Brammer, M. J. (2002). Speechreading circuits in people born deaf. *Neuropsychologia, 40*(7), 801–807.

Magnotti, J. F., & Beauchamp, M. S. (2015). The noisy encoding of disparity model of the McGurk effect. *Psychonomic Bulletin & Review*, *22*(3), 701–709.

Magnotti, J. F., & Beauchamp, M. S. (2017). A causal inference model explains perception of the McGurk effect and other incongruent audiovisual speech. *PLoS Computational Biology*, *13*(2), e1005229.

Magnotti, J. F., Mallick, D. B., & Beauchamp, M. S. (2018). Reducing playback rate of audiovisual speech leads to a

surprising decrease in the McGurk effect. *Multisensory Research*, *31*(1-2), 19–38.

Magnotti, J. F., Mallick, D. B., Feng, G., Zhou, B., Zhou, W., & Beauchamp, M. S. (2015). Similar frequency of the McGurk effect in large samples of native Mandarin Chinese and American English speakers. *Experimental Brain Research*, *233*(9), 2581–2586.

Mallick, D. B., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review*, *22*(5), 1299–1307.

Marques, L. M., Lapenta, O. M., Costa, T. L., & Boggio, P. S. (2016). Multisensory integration processes underlying speech perception as revealed by the McGurk illusion. *Language, Cognition and Neuroscience*, *31*(9), 1115–1129.

Marques, L. M., Lapenta, O. M., Merabet, L. B., Bolognini, N., & Boggio, P. S. (2014). Tuning and disrupting the brain-modulating the McGurk illusion with electrical stimulation. *Frontiers in Human Neuroscience, 8*, 533.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

Miller, L. M., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *The Journal of Neuroscience*, *25*(25), 5884–5893.

Moro, S. S., & Steeves, J. K. E. (2018). Audiovisual plasticity following early abnormal visual experience: Reduced McGurk effect in people with one eye. *Neuroscience Letters, 672*, 103–107.

Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception, & Psychophysics*, *58*(3), 351–362.

Munhall, K. G., ten Hove, M. W., Brammer, M., & Paré, M. (2009). Audiovisual integration of speech in a bistable illusion. *Current Biology*, *19*(9), 735–739.

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk Eeffect, a multisensory speech illusion. *NeuroImage*, *59*(1), 781–787.

Nath, A. R., Fava, E. E., & Beauchamp, M. S. (2011). Neural correlates of interindividual differences in children's audiovisual speech perception. *The Journal of Neuroscience*, *31*(39), 13963–13971.

Olasagasti, I., Bouton, S., & Giraud, A. L. (2015). Prediction across sensory modalities: A neurocomputational model of the McGurk effect. *Cortex*, *68*, 61–75.

Palmer, T. D., & Ramsey, A. K. (2012). The function of consciousness in multisensory integration. *Cognition*, *125*(3), 353–364.

Paré, M., Richler, R. C., ten Hove, M., & Munhall, K. G. (2003). Gaze behavior in audiovisual speech perception: The influence of ocular fixations on the McGurk effect. *Perception, & Psychophysics*, *65*(4), 553–567.

Proverbio, A. M., Massetti, G., Rizzi, E., & Zani, A. (2016). Skilled musicians are not subject to the McGurk effect. *Scientific Reports*, *6*, 30423.

Quinto, L., Thompson, W. F., Russo, F. A., & Trehub, S. E. (2010). A comparison of the McGurk effect for spoken and sung syllables. *Attention, Perception, & Psychophysics*, *72*(6), 1450–1454.

Romero, Y. R., Senkowski, D., & Keil, J. (2015). Early and late beta-band power reflect audiovisual perception in the McGurk illusion. *Journal of Neurophysiology*, *113*(7), 2342–2350.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics*, *59*(3), 347–357.

Rosenblum, L. D., Yakel, D. A., & Green, K. P. (2000). Face and mouth inversion effects on visual and audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 806–819.

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, *17*(5), 1147–1153.

Rouger, J., Fraysse, B., Deguine, O., & Barone, P. (2008). McGurk effects in cochlear-implanted deaf subjects. *Brain Research*, *1188*(1), 87–99.

Saint-Amour, D., De Sanctis, P., Molholma, S., Ritter, W., & Foxe, J. J. (2007). Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. *Neuropsychologia*, *45*(3), 587–597.

Samuel, A. G. (2011). Speech perception. *Annual Review of Psychology*, *62*(1), 49–72.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics*, *59*(1), 73–80.

Sekiyama, K., Soshi, T., & Sakamoto, S. (2014). Enhanced audiovisual integration with aging in speech perception: A heightened McGurk effect in older adults. *Frontiers in Psychology*, *5*, 323.

Sekiyama, K., & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, *21*(4), 427–444.

Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the McGurk–MacDonald illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 580–587.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*, 255–266.

Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1517–1529.

Strand, J., Cooperman, A., Rowe, J., & Simenstad, A. (2014). Individual differences in susceptibility to the McGurk effect: Links with lipreading and detecting audiovisual incongruity. *Journal of Speech Language and Hearing Research*, *57*(6), 2322–2331.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philosophical Transactions: Biological Sciences, 335*(1273), 71–78.

Thomas, S. M., & Jordan, T. R. (2002). Determining the influence of Gaussian blurring on inversion effects with talking faces. *Perception & Psychophysics*, *64*(6), 932–944.

Thomas, S. M., & Jordan, T. R. (2004). Contributions of oral and extraoral facial movement to visual and audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(5), 873–888.

Tiippana, K. (2014). What is the McGurk effect? *Frontiers in Psychology*, *5*, 725.

Tiippana, K., Andersen, T. S., & Sams, M. (2004). Visual attention modulates audiovisual speech perception. *European Journal of Cognitive Psychology*, *16*(3), 457–472.

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*(8), 1096–1101.

Ujiie, Y., Asai, T., & Wakabayashi, A. (2015). The relationship between level of autistic traits and local bias in the context of the McGurk effect. *Frontiers in Psychology*, *6*, 891.

Ujiie, Y., Asai, T., & Wakabayashi, A. (2018). Individual differences and the effect of face configuration information in the McGurk effect. *Experimental Brain Research, 236*(4), 973–986.

Van Engen, K. J., Xie, Z., & Chandrasekaran, B. (2017). Audiovisual sentence recognition not predicted by susceptibility to the McGurk effect. *Attention Perception & Psychophysics*, *79*(2), 396–403.

Walker, S., Bruce, V., & O'Malley, C. (1995). Facial identity and facial speech processing: Familiar faces and voices in the McGurk effect. *Perception & Psychophysics*, *57*(8), 1124–1133.

Wilson, A. H., Alsius, A., Paré, M., & Munhall, K. G. (2016). Spatial frequency requirements and gaze strategy in visual-only and audiovisual speech perception. *Journal of Speech, Language, and Hearing Research*, *59*(4), 601–615.

Zhu, L. L., & Beauchamp, M. S. (2017). Mouth and voice: A relationship between visual and auditory preference in the human superior temporal sulcus. *The Journal of Neuroscience*, *37*(10), 2697–2708.

The influential factors and neural mechanisms of McGurk effect

LUO Xiaoxiao[1]; KANG Guanlan[1]; ZHOU Xiaolin[1,2,3,4]

([1] *School of Psychological and Cognitive Sciences, Peking University, Beijing,* 100871, *China*)

([2] *Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing* 100871, *China*)

([3] *PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing* 100871, *China*)

([4] *Institute of Psychological and Brain Sciences, Zhejiang Normal University, Jinhua* 321004, *China*)

Abstract: The McGurk effect is a typical audiovisual integration phenomenon, influenced by characteristics of physical stimuli, attentional allocation, the extent that individuals rely on visual or auditory information in processing, the ability of audiovisual integration, and language/culture differences. Key visual information that leads to the McGurk effect is mainly extracted from the mouth area of the talker. The McGurk effect implicates both audiovisual integration (which occurs in the early processing stage and is related to the activation of superior temporal cortex) and the conflict of the incongruent audiovisual stimuli (which occurs in the late processing stage and is related to the activation of inferior frontal cortex). Future studies should further investigate the influence of social factors on the McGurk effect, pay attention to the relationship between unimodal information processing and audiovisual integration in the McGurk effect, and explore the neural mechanisms of McGurk effect with computational modeling.

Key words: McGurk effect; audiovisual speech perception; audiovisual integration; multisensory integration